LSHSS

Research Article

Investigating Listening Effort in Classrooms for 5- to 7-Year-Old Children

**Nicola Prodi,[a] Chiara Visentin,[a] Alessandro Peretti,[b]
Jacopo Griguolo,[c] and Giovanni Battista Bartolucci[d]**

AQ1 **Purpose:** This study aims to investigate the listening effort made by young children in real classrooms during a prolonged speech reception task in the presence of background noise.

Method: The experiment was proposed to 117 typically developing kindergarten and primary school pupils, aged 5–7 years old. An ecological experimental approach was followed, and speech-in-noise tests were presented in the classrooms to groups made up of the whole class. The speech material of the Word Intelligibility by Picture Identification Test in the Italian language (Arslan, Genovese, Orzan, & Turrini, 1997) was presented in 2 listening conditions (quiet classroom [no noise added] and working classroom [with stationary noise]) and was repeated twice during the experiment. Data on the number of correctly recognized words and the single-task response time (RT) were collected; the quantity of the latter was considered informative on listening effort.

Results: It was found that when background noise was present, the pupils' performance decreased, and greater RTs were required compared to the "quiet classroom" condition. When the RTs were analyzed over the course of the experiment, there were no changes in the quiet condition, whereas in the working classroom, a significant increase was found for the 6- and 7-year-old pupils. On the contrary, the youngest pupils (5-year-olds) showed a decrease in the RT results over the test repetitions.

Conclusions: The RT measured with a single-task paradigm was found to be a viable approach for investigating the listening effort in 6- to 7-year-old pupils. For this age range, the metric was sensitive to changes both in the listening conditions and within the same listening condition across the time of exposure. More research is needed to assess the feasibility of the experimental paradigm with the 5-year-old children.

Noise at school can be harmful to the speech perception and listening comprehension of children (Crandell & Smaldino, 2000; Klatte, Bergström, & Lachmann, 2013), so much so that the sound environment in classrooms can be inappropriate for learning (Anderson, 2004; Nelson & Soli, 2000; Valente, Plevinsky, France, Heinrichs-Graham, & Lewis, 2012). In particular, over the years, it has been shown that children's performance in diverse tasks is affected by noise and that, as a consequence, their academic achievements can be mined (Shield & Dockrell, 2003, 2008). Different trends can be outlined according to age, mainly because younger pupils are more prone to errors due to a less mature auditory system and cognitive development (Leibold, Bonino, & Buss, 2016; Prodi, Visentin, & Feletti, 2013). A developmental effect was also observed when the background noise consisted of intelligible talkers (speech-in-speech masking), greater than that observed in the presence of energetic masking alone; the effect is typically attributed to immature segregation and/or selective attention (Leibold et al., 2016). With regard to speech reception, it is a basic necessity for communication and a prerequisite for more structured tasks in the classroom. Its accuracy is differentially impacted by different types of noise (Astolfi, Bottalico, & Barbato, 2012; Leibold et al., 2016; Prodi et al., 2013; Sato & Bradley, 2008), even though the acoustical conditions are characterized by the same objective indicators, such as the speech transmission index (STI; International Electrotechnical Commission, 2011).

In any case, the problematic experience of listening in the classroom is not resolved by controlling speech reception

---

[a]Department of Engineering, University of Ferrara, Italy
[b]Postgraduate School in Occupational Medicine, University of Padua, Italy
[c]Jacopo Griguolo Engineering: Acoustic, Measurements and Design, Rovigo, Italy
[d]Department of Cardiothoracic and Vascular Sciences, University of Padua, Italy
AQ2 Correspondence to Nicola Prodi: nicola.prodi@unife.it

Editor-in-Chief: Shelley Gray
Editor: Dawna Lewis
Received March 6, 2018
Revision received July 1, 2018
Accepted September 19, 2018
https://doi.org/10.1044/2018_LSHSS-18-0039

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

accuracy alone, which ensures that a high proportion of the utterances are correctly received by the children. In fact, the dimension of the listening-related effort is also crucial in the assessment of communication in classrooms. Several factors may require that an increased amount of cognitive resources is called for during speech reception; these factors can either be related to the signal generation (e.g., speakers' pronunciation, intonation) and to the transmission channel (e.g., suboptimal acoustic conditions, type of noise) or to the listener (e.g., age, hearing loss, language proficiency). As a result, an outcome of effort is often elicited (McGarrigle et al., 2014), especially so in the presence of sustained listening demands, such as during lessons. *Listening-related effort* can be defined as a specific form of a more general construct known as "mental effort"; this involves listening tasks and embeds both the attentional and capacity allocation policies of the listeners (Pichora-Fuller et al., 2016). The qualification of effortful listening has been proposed with diverse paradigms pertaining to physiological, behavioral (task performance–based), and subjective (self-report) domains (Klink, Schulte, & Meis, 2012a, 2012b; McGarrigle et al., 2014; Pichora-Fuller et al., 2016).

With regard to behavioral measures, they can follow two different approaches, termed *dual-task* and *single-task*. In dual-task experiments, a speech-processing primary task is paired with a secondary task, the performance of which is monitored (Gagné, Besser, & Lemke, 2017). Because of the limited capacity hypothesis (Kahnemann, 1973), the decrement of performance in the secondary task is assumed as proof of the increase in listening effort. When applied in studies involving children, this approach highlighted the listening-related effort for children with hearing loss (Hicks & Tharpe, 2002) and its increase with the lowering of speech-to-noise ratios (SNRs) for pupils aged 9–12 years (Howard, Munro, & Plack, 2010). Concerns about the reliability of dual-task paradigms for children have been raised (Choi, Lotto, Lewis, Hoover, & Stelmachowicz, 2008) because of the difficulty of pupils to control the allocation of resources according to instructions, and it is plausible that younger children would have greater difficulties in this respect. Thus, single-task auditory experiments have been considered, which rely on the measurement of verbal response time (RT). The underlying assumption is that single-task RT is effective in tracing working memory operations involved with speech processing in children (Cowan et al., 2003). Consistently with studies in adults (Houben, van Doorn-Bierman, & Dreschler, 2013; Pals, Sarampalis, van Rijn, & Başkent, 2015; van den Tillan-Haverkate, de Ronde-Brons, Dreschler, & Houben, 2017), a longer RT would indicate an increase in cognitive load and hence in processing effort. Verbal RT within a single-task paradigm was used to assess listening effort in school children. For instance, Gustafson, McCreery, Hoover, Kopun, and Stelmachowicz (2014) evaluated the impact of noise reduction (recorded through amplification) in a panel of pupils with normal hearing aged 7–12 years. Moreover, AQ3 Choi, Lotto, Lewis, Hoover, and Stelmachowicz (2016) traced the increase in effort with the decrease in SNR that occurs both for children with normal hearing (aged 5–12 years) and for children (aged 8–12 years) with mild hearing loss. It is important to notice that single-task experiments with retrieval of RT also have shortcomings. For instance, the attention of participants might depend on the difficulty of the task itself, so that the relationship between mental effort and RT could change, depending on the experimental conditions (Bess & Hornsby, 2014; McGarrigle et al., 2014). Should this be the case, the effects of arousal might mask the effect of the factors under investigation on the RT measurement.

On the other hand, the single-task paradigm allows the approach to be implemented within the framework of field experiments inside real-life classrooms. Technological advances have provided means such as personal response devices (Vickers et al., 2013) or smartphone applications (Prodi, Visentin, & Bellettini, 2012) that allow an entire class of pupils to be tested at once, also preserving self- and mutual perceptions, as experienced during lessons. Presenting speech-in-noise tests inside real-life scenarios is a valuable ecological opportunity; in this context, the ecological validity is intended as "the degree to which a study accurately represents the conditions under which an effect occurs in the real world" (Reis, 2012). To date, children's listening effort has mainly been investigated within the framework of laboratory experiments, with only a few studies dealing with the topic in real-life scenarios. For instance, Mealings, Buchholz, Demuth, and Dillon (2015) studied open-plan classrooms to outline the simultaneous detrimental effects on the accuracy and speed of processing due to the open-plan design, and Prodi et al. (2013) provided a ranking of the impact of noise types on speech reception in primary school classrooms, with reference to speech intelligibility (SI) and manual RT.

Then, even though the existing literature on single-task RT and children is rapidly increasing, at present, only a few studies have focused on experiments in real-life classrooms, dealing with the combined effect of reverberation and background noise. Furthermore, to the best of the authors' knowledge, only two studies (Lewis et al., 2016; Mealings et al., 2015) tested the single-task paradigm with reference to very young children (5 years old). Both found an increase in RT as the listening conditions worsened.

In this study, speech-in-noise tests were presented to young children with normal hearing in their classrooms. Still within the framework of a controlled playback of the stimuli, the in situ presentation was chosen in order to preserve the children's spatial awareness during the experiment, and the pupils performed a speech reception task in a complex listening situation, together with their classmates. The experimental paradigm was intended to present the children with a more ecological environmental condition, thus providing stronger face validity than laboratory testing.

The study extends the analysis outlined by Prodi and Visentin (2015) to a group of younger children. In that study, a single-task paradigm was used to retrieve RT data for 8- to 10-year-old pupils, and the same ecological paradigm as depicted above (a speech recognition task for the entire class at once, inside their classroom) was followed. The results highlighted the complexity of the interaction between noise type, pupils' age, and acoustical conditions in a realistic classroom setup. In particular, when considering internal activity noise as a masker, a developmental trend in the RT results was found, with the older pupils showing the fastest RT results especially in the most favorable acoustic conditions. Differently, McCreery and Stelmachowicz (2013), in a word repetition task with 6- to 12-year-old children, found that age was not correlated with mean verbal processing time. A complex pattern was found by Lewis et al. (2016), with the onset time decreasing as the age increased; the effect was mediated by the interaction between stimulus complexity and SNR (longer RTs with decreasing SNR and more complex stimuli; i.e., sentences). The differences in the results of these studies indicated how the stimulus complexity and the listening conditions (namely, the spectral and temporal characteristics of the background noise) used in the experiment might influence the outcomes.

In this study, the relationship between the children's age and the acoustic environment was explored, aiming to extend the findings of Prodi and Visentin (2015) to a group of younger children and to reproduce the increase in RT as the listening conditions worsened, as found by Mealings et al. (2015) and Lewis et al. (2016). Furthermore, the feasibility of doing a single-task paradigm to measure listening effort was tested. The latter goal was motivated by the necessity to account for the influence of room acoustics (and its modifications) on the younger children's performance, not only with regard to speech reception but also with regard to cognitive level. Indeed, whereas the traditional normative approach to classroom design relies solely on SI results, it is believed that the additional information on the processing effort provided by the single-task RT would allow the effect of room acoustics to be better discerned beyond performance accuracy alone.

SI and single-task RT were then investigated, and three research questions were formulated:

1. When the children are tested inside their classroom, how does the RT vary in the presence of background noise?
2. Is there a relationship between the noisy environment of the classrooms and the occurrence of effort?
3. Could the RT implemented in a single-task paradigm be used as a proxy of listening effort for 5- to 7-year-old children with normal hearing and typical development tested in their classrooms?

Method

The study focused on the assessment of speech recognition and listening effort of young children, measured during group sessions in their classrooms. The children were asked to complete a word recognition task implemented by using disyllabic target words presented in two listening conditions (ambient noise and steady-state speech-shaped noise). The task was presented in a closed-set format by using personal touch screen devices assigned to each child.

Description of the Classrooms

The study took place in Padua (Italy) in the second half of the school year; it involved a kindergarten and a primary school located in the city center, far from main roads and surrounded by the schoolyard.

One classroom was selected in each school to be used as a laboratory for the speech-in-noise tests and the acoustical measurements. In the kindergarten, where only one class of pupils performed the experiment, their classroom was selected as the laboratory. In the primary school, the classroom of a group of Grade 1 children was selected, and the other classes went there for the experiment on a rotational basis. In this case, it was ensured that the selected space was similar for geometrical and acoustical characteristics to the remaining classrooms of the school. In the kindergarten, the laboratory classroom was located on the ground floor, whereas in the primary school, it was on the second floor; both overlooked the schoolyard. The classrooms were box-shaped; the dimensions were 6.9 × 9.9 × 4.9 m in the kindergarten and 6.0 × 6.6 × 4.2 m in the primary school, resulting in a volume of 337 and 166 m^3, respectively. The classrooms had flat surfaces, with a plaster finishing of the lateral surfaces and ceramic tile flooring. Here, it is important to notice that, whereas in the primary school the ceiling was acoustically treated with sound absorbing panels, no acoustical treatment was present in the kindergarten. Windows were present on one side only, and at least another lateral wall bordered onto a corridor. The classrooms were designed for a maximum of 25 pupils. During the experiment, the classrooms were set up as for regular lessons with wooden desks and chairs, closets, and shelves.

Classrooms Setup and Measurements of the Room Acoustics

For the experiment, the classrooms were set up as follows. A Gras 44AB mouth simulator was placed at the teacher's desk, 1.5 m high from the ground (assumed as the mouth height of a standing teacher); it was oriented toward the audience and used to deliver the speech signal for the speech-in-noise tests or the test signal for the measurement of the acoustical parameters. An additional loudspeaker (Quested F11) was used to playback interfering noise during the experiment. It was placed on the floor, to the side of the mouth simulator, facing at the opposite direction of the audience to minimize the impact of the direct sound. Two measurement positions were defined in the area where the children were seated during the experiment. Ideally, the seating area was divided into four quadrants (front-right, front-left, back-right, back-left) and monaural microphones were placed at the center of two opposite quadrants (front-right and back-left). The distances between the speech source and the receiving positions were 3.0 and

4.4 m in the kindergarten classroom and 2.6 and 3.6 m in the primary school classroom. The plans and layouts of the F1 two laboratory classrooms are shown in Figure 1.

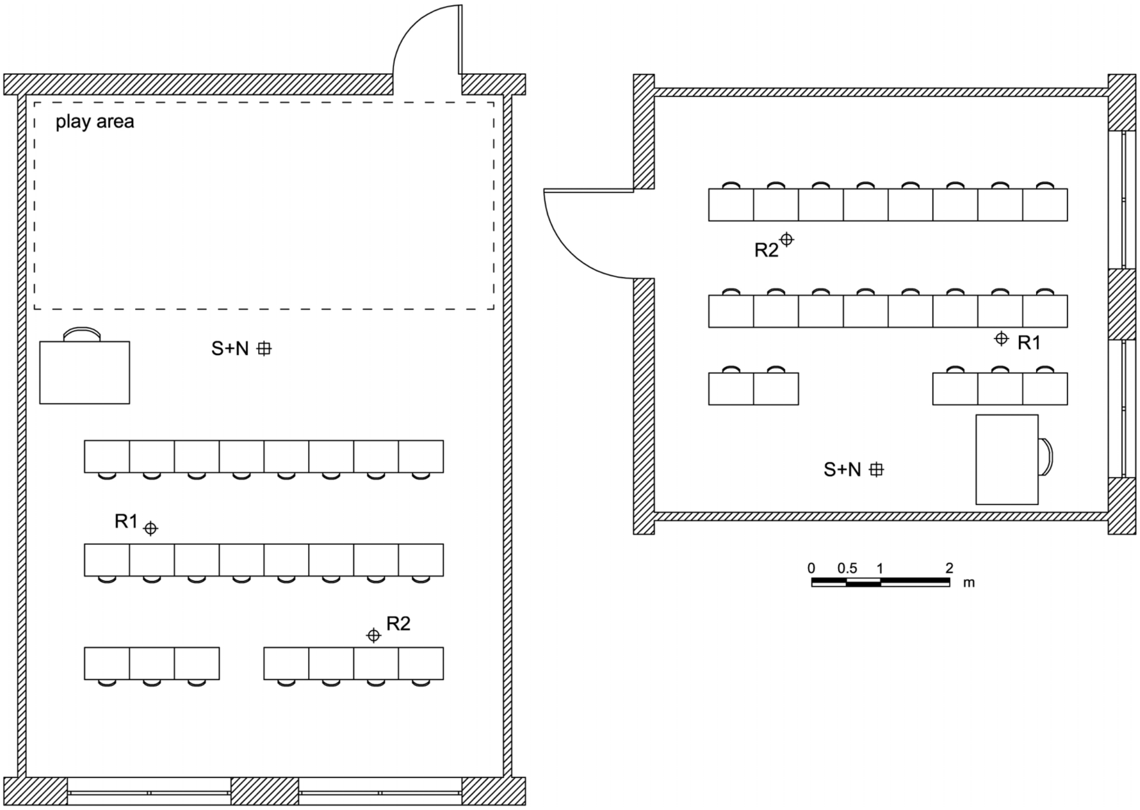
The objective acoustic characterization of the two laboratory classrooms was achieved at the listening positions by using two omnidirectional B&K Type 4189 0.5-in. microphones, set at a height of 1.1 m, taken as the height of a seated child's ears. Furthermore, the measurement setup included a B&K Type 5935 signal conditioner, a B&K Type 4231 calibrator, an RME Fireface UC full-duplex sound card, and a laptop. The signal playback and recording and the following acoustic elaborations were managed by means of the Aurora suite in the Adobe Audition package. In particular, in the classrooms in occupied conditions, the octave band L_{eq} was sampled over a 30-s interval for the speech signal, and the background noises (ambient noise and interfering noise from the loudspeaker) were reproduced at the same levels as during the listening tests. The SNR at the two listening positions were calculated from the measurements. Then, a sine sweep over the 20-Hz to 20-kHz frequency interval was played back from the speech source and recorded in order to derive the impulse responses at the listening positions. The impulse responses were used to calculate the reverberation time, defined as the time it takes the sound to decrease by 60 dB. For this purpose, the indicator T_{20} was considered (International Organization of Standardization, 2008, 2009). Furthermore, the STI was also calculated with the indirect method (International Electrotechnical Commission, 2011).

Participants

The speech-in-noise tests were presented to 117 kindergarten and primary school children aged 5–7 years. The group of 5-year-old children (hereinafter 5Y) was composed of 23 pupils from one class attending their last year of kindergarten. The group of older children was instead composed of 94 pupils attending their first or second grade at primary school, thus being either 6 or 7 years old; from now on, the groups will be referred to as 6Y (three classes, 49 pupils) and 7Y (two classes, 45 pupils). Written consent forms granting permission for the children to participate in the experiment were obtained from the parents.

After the experiment, the teachers were asked to provide details on children with certified intellectual or behavioral disabilities, children with certified hearing impairment, and children having lived in Italy for less than 1 year at the date of the experiment. Overall, there were four children with such characteristics (two in each school) who were excluded from the subsequent data analysis. Furthermore, the

Figure 1. Setup of the kindergarten (left) and the primary school (right) classrooms during the experiment. The signal (S1) and noise (S2) sources were located close to the teacher's desk; the acoustical measurements were performed at positions R1 and R2.



children's linguistic background was investigated using a questionnaire, compiled by the teachers, referring to the information provided by parents at the child's enrollment. Specifically, it contained questions on the child's place of birth, the years that he or she has lived in Italy, the parents' mother tongue(s), and the language(s) spoken at home with parents and siblings. The answers indicated that the two schools were mainly attended by pupils of Italian mother tongue, which represented 88.5% of the students participating in the experiment. Because of the impact language experience may have on the speech reception task, only children speaking Italian as their primary language were considered in the data analysis. The final T1 group of participants is detailed in Table 1.

Stimuli

Speech reception was assessed using the Word Intelligibility by Picture Identification Test (Arslan, Genovese, Orzan, & Turrini, 1997), in its implementation in the Italian AQ4 language named *TIPI*. The TIPI is a closed-set, picture-pointing test, which bases on disyllabic word, all of them belonging to the everyday vocabulary of a 4-year-old child. The test *corpus* is composed of 96 nouns that are graphically represented through simple handwritten, pencil-colored pictures. Fifty words out of 96 are the target items, organized in pairs of phonetically similar items; the words within a "minimal pair" differ by one consonant (e.g., /ˈ*vino*/ and /ˈ*pino*/, /ˈ*barba*/ and /ˈ*barka*/). The remaining 46 items are distracters. The words are organized into 50 groups of six items, including the target item, its phonetically similar pair, and four distracters (two items with the same vowel as the target and two items not sharing any phonetical features with the target).

The TIPI items are not phonetically balanced, whereas they are designed to ensure the same probability of word occurrence within the children's vocabulary. The test is commonly used in the field of audiology for testing pupils aged 4 years and older. In fact, the closed-set format based on pictures instead of words allows the test to be used with young children not having yet developed reading skills, and the test material is conceived to overcome the issues related to their still limited linguistic competences and attentional capacities.

For the experiment, sentences composed by a carrier phrase and a target word belonging to the TIPI *corpus* were created (e.g., "Ora diremo la parola *barba*"—which means "Now we will say the word *beard*"). The sentences were recorded by an adult, native Italian, female speaker; she was instructed to speak at a conversational rate, maintaining natural intonation and avoiding any emphasis on the final, target word. The recordings took place in a silent room using a B&K Type 4189 0.5-in. microphone placed about 15 cm from the speaker's mouth and routed to a B&K Type 5935 signal conditioner. The digital recordings had a sampling frequency of 44.1 kHz. All of the sentences were filtered to match the long-term spectrum of a female speaker indicated by the IEC60268-16 standard (International Electrotechnical Commission, 2011) and set to the same root-mean-square value. The recordings were then organized into five lists of 10 words each, using a pseudo-randomized procedure to ensure that the two words of a "minimal pair" never occurred within the same list. The equivalence of the test lists (as concerns the intelligibility results) was assessed in a pilot laboratory study with 6-year-old pupils; the lists were found to be equivalent. For the experiment, one list was randomly selected and used in the training phase, whereas the remaining four were used in the speech-in-noise tests.

Listening Conditions

Two listening conditions were set, named *quiet classroom* and *working classroom.* In both conditions, the speech signal was calibrated to a level of 63 dB(A), measured at 1 m in front of the mouth simulator using a B&K 4165 0.5-in. microphone, a B&K 2639 preamplifier, a B&K 5935 signal conditioner, and a B&K 4231 sound-level calibrator. This corresponds to a speaker talking with intermediate vocal effort between "normal" and "raised" (International Organization of Standardization, 2003). Bottalico and Astolfi (2012) investigated the vocal effort of primary school teachers and found that the average value over a working day of the mean sound pressure level of the voiced speech measured at 1 m from a female speaker's mouth was 62.1 dB(A).

In the "quiet classroom" condition, the pupils performed the task in the actual ambient noise of the classroom, primarily made up of noises coming from the adjacent classrooms, where pupils were engaged in quiet activities (teaching time or individual work). On the contrary, in the "working classroom" condition, a long-term speech-shaped

Table 1. Characteristics of the children participating in the experiment.

School	Grade	Group	No. of participants	No. of male/female	Age [year;months] (*SD*)
Kindergarten		5Y	18	10/8	5;10 (2)
Primary school	Grade 1	6Y	43	18/25	6;10 (4)
	Grade 2	7Y	39	19/20	7;9 (4)
All			100	47/53	7;0 (9)

Note. For each school, the number of participants considered in the data analysis is detailed, further divided between male and female. The mean age of each group is indicated in (years;months), together with the corresponding standard deviation (*SD*).

noise (LTSS) was played back by the additional loudspeaker producing an energetic masking of the signal. The noise was obtained starting from a steady-state pink noise, which was spectrally shaped in octave bands to match the long-term spectrum of female speech (International Electrotechnical Commission, 2011). The noise level was calibrated similarly to the speech signal to obtain an SNR of +3 dB at 1 m in front of the mouth simulator and of 0 dB in the seating area in occupied conditions.

An objective description of the listening conditions at the receiving positions was achieved by using the STI, which describes the combined effect of steady background noise and reverberation on the transmission quality of the speech signal. For this purpose, the long-term levels of the ambient noise, the speech signal, and the LTSS noise (reproduced as proposed during the tests) were measured at the end of the experiment, with the classroom still in occupied conditions.

When several classes took the experiment in the same classroom (i.e., in the primary school), the acoustic measurements were replicated for each group, and the final values were obtained as the average of the repetitions; only minimal discrepancies were found between the replications. The average measured values of the two receiving positions are reported in Table 2 for each school. The usage of spatially averaged values is justified by the fact that, for all the acoustical parameters, the values measured at the two receivers always differed for a quantity smaller than the corresponding "just noticeable difference": 5% for the reverberation time and 1 dB for the sound pressure level (International Organization of Standardization, 2009). Concerning the STI, the difference was always smaller than the rating interval of 0.04 defined in the IEC 60268-16 standard (International Electrotechnical Commission, 2011). T2

When ambient noise alone was considered ("quiet classroom" condition), the measured SNRs were +16 dB (kindergarten) and +15.7 dB (primary school), both being higher than the limit of +15 dB defined by Picard and Bradley (2001) as the acceptable value for speech communication in classrooms. Nonetheless, because of the difference in the room acoustics, the resulting sound environment varied considerably between the two laboratory classrooms and the only primary school classroom, which had undergone an acoustical treatment of the ceiling, met the target value (STI = 0.6) defined for classrooms by the Italian standard UNI11367-Annex C (Ente Nazionale Italiano di Unificazione, 2010). The measured STI values were 0.73 and 0.56, thus corresponding to intelligibility rated as "good" in the primary school and only "fair" in the kindergarten (International Organization of Standardization, 2003). This inherent difference in the classroom conditions prevents a direct comparison between the 5Y and the 6Y and 7Y whose performance will be separately addressed in the following analysis. The "quiet classroom" condition, with ambient noise alone, is the most comfortable classroom condition that the pupils could experience within their respective classrooms during lessons; hence, it was assumed as the baseline, against which the condition with additional stationary noise was compared.

This latter condition was set to reproduce a working classroom, where the pupils carry out educational activities with their teachers. The sound level at the audience corresponds to the average level measured in primary classrooms by Shield and Dockrell (2004) when the children are engaged in the quietest activities (e.g., reading or doing a test). The resulting STI values, reported in the last row of Table 2, correspond to intelligibility rated as "poor" in the kindergarten and "fair" in the primary school (International Organization of Standardization, 2003).

Procedure

The experiment took place in groups made up of the whole class, over the course of one morning for each school. In the primary school, the five classes of pupils involved in the experiment went to the laboratory classroom on a rotational basis to perform the speech-in-noise tests. Upon entering the laboratory classroom, each child was given a touch screen handset, to be used for response selection, and was randomly assigned a seating position. Children sat at desks, which were arranged in regular lines, with all positions facing the front of the classroom and, therefore, the

Table 2. Listening conditions within the two classrooms during the experiment: reverberation times ($T_{20,\mathrm{mid}}$, average over the 500- to 2000-Hz octave bands), A-weighted sound pressure levels for the speech signal, the ambient noise and the additional stationary noise, and resulting speech transmission index (STI) values for the conditions of "quiet classroom" and "working classroom."

Acoustic parameters	Kindergarten (5Y group)	Primary school (6Y and 7Y groups)
$T_{20,\mathrm{mid}}$ [s]	1.31	0.54
$L_{A,\mathrm{eq}}$ [dB(A)]		
Speech	56.9	57.8
Ambient noise	40.9	42.1
LTSS noise	56.1	57.5
STI		
Quiet classroom	0.56	0.73
Working classroom	0.42	0.53

Note. The reported data are averaged across the two receiving positions within each classroom. LTSS = long-term speech-shaped noise.

loudspeakers. Each child was also assigned a unique identification code to provide a correct coupling between listening position, test device, and user. The usage of an identification code also ensured anonymity when handling the results.

Before the test session, the experiment was presented to the class. The TIPI pictures were shown to the children, indicating and repeating together the corresponding target words. The task was carefully explained and practiced with examples in order to familiarize the children with the touch screen handset and the data collection system. In fact, a wireless test bench was used to manage the experiment (Prodi et al., 2013); the server application running on a laptop simultaneously controlled the audio rendering, the presentation of the pictures on the touch screen handsets, and the collection of the responses.

The children listened to a target word embedded in the carrier phrase; when the background noise was played back, it started approximately 1,000 ms before the carrier phrase and ended simultaneously to the speech signal. At the offset of the audio playback, six pictures were displayed on the touch screen, and the children were instructed to select the picture that corresponded to the word they heard. They were allowed a maximum of 20 s to respond, and only once all participants had responded (or reached the timeout) was the next target word automatically presented. No feedback was given to the children regarding correct answers.

A practice session was initially proposed, where a 10-word test list was presented to the children, first in quiet conditions and then with background noise. The aim of the practice was twofold: first, to ensure that the children were familiar with the listening conditions proposed in the experiment and especially with the type and level of noise played back during the "working classroom" condition and, second, to ensure that all the children understood the assignment and were able to use the touch screen device correctly. The practice session was presented once for primary school children; it was repeated twice in the kindergarten, in order to be confident that all children were able to give a response within the timeout.

Afterward, they completed four listening tests (2 test lists × 2 listening conditions); each test comprised 10 target words. The listening conditions were thus proposed two times each (referred to as repetitions R1 and R2 in the following) and alternated during the presentation, blocking the order within the repetitions. The presentation order could be A (quiet–working–quiet–working classroom) or B (working–quiet–working–quiet classroom). In the primary school (where five classes were tested), the presentation order of the listening conditions was counterbalanced across the classes in order to minimize the influence of sequential and learning effects; furthermore, the test lists were pseudorandomized, as to avoid the same test list always being coupled to the same listening condition. In the kindergarten, where only one class was tested, the listening conditions were presented in the A order.

The experiment duration was carefully planned in order to remain within acceptable time limits, lasting a maximum of 30 min. The targeted duration was derived from the experimental paradigm of Prodi and Visentin (2015) where, for the older children aged 8–10 years, the performance was examined over a longer 45-min period. A shorter duration was then targeted for younger pupils, accounting for their less mature cognitive development and shorter attentional capabilities. The time constraint also determined the maximum number of 20 target words that could be presented in each listening condition; however, the number of stimuli per condition was comparable with other literature studies with young children (Gustafson et al., 2014; Lewis et al., 2016).

The entire experimental session, from the practice to the final listening test, lasted approximately 30 min for the 5Y children, comprising 15 min for the practice and 15 min for the tests. Over the four test repetitions, the 20-s timeout was reached in seven occasions (three times in the "quiet classroom" condition and four times in the "working classroom" condition). As concerns the primary school, the experiment lasted on average 20 min (10 min for the practice and 10 min for the tests); the 20-s timeout was reached only twice.

An outline of the experiment design is reported in Figure 2. The children were instructed to pay attention to the task and asked to respond as accurately as possible. They were not informed that the RT data were also acquired during the experiment, nor were they given any recommendations to respond as quickly possible. However, during the experiment, the teachers and two members of the research team monitored the pupils to ensure that the responses were provided in a timely manner. **F2**

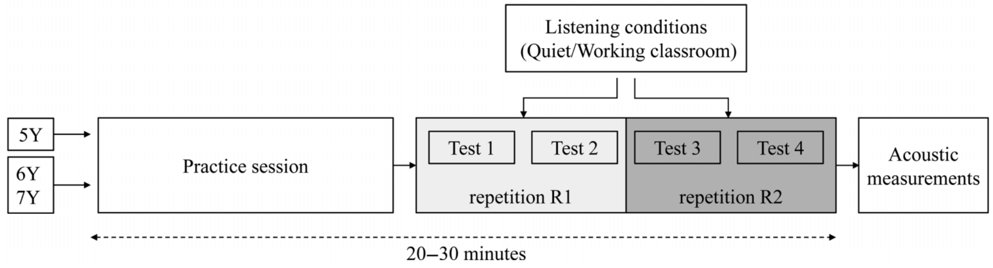
The data collected during the experiment were the picture scores (correct/incorrect/phonetically similar) and the RTs, defined as the time elapsed between the audio offset and the response selection on the touch screen.

Statistical Analysis

A generalized linear mixed model (GLMM) was used for the statistical analyses, using the software R (R Core Team, 2017) and the lme4 package (Bates, Maechler, Bolker, & Walker, 2015). An $\alpha = .05$ significance level was always assumed. The statistical method was selected due to the possibility of dealing at once with response variable distributions departing from the normal distribution and repeated measurements of the participants. **AQ5**

Different models have been built for the two response variables. A GLMM with a binomial distribution was used for the SI results; the response variable in the model was coded with a binary score (0/1 corresponding to a wrong/correct response). For the analysis, phonetically similar responses were considered as incorrect. Regarding the RT data, they were modeled assuming a Gamma distribution with a log link function (Baayen & Milin, 2010; Lo & Andrews, 2015), selected to reproduce the characteristics of raw RT data, that is a positively skewed continuous distribution raising rapidly on the left and having a long positive tail on the right (Whelan, 2008). Prior to the analyses, an a priori screening was performed on the RT results and

Figure 2. Outline of the experimental design, detailing the groups of participants (the 5Y and the 6Y and 7Y) and the temporal flow from the practice session to the final acoustical measurements in the occupied laboratory classrooms. The order of listening conditions (quiet classroom, working classroom) was blocked within the repetitions (R1 and R2) and counterbalanced across the six classes of children performing the experiment.



excessively long data, possibly due to participants' inattention, were removed. A cutoff of 10 s was set, beyond which the RT results were discarded and considered as missing data; altogether, 34 RTs were removed (1.2% of the data set).

For all statistical analyses, model selection was based on a forward procedure using a likelihood ratio test. The statistical assumptions of the final models have been verified by checking the normality of the random effect terms and the residuals (Everitt & Hothorn, 2010). In case of statistically significant effects of the main factors or of the interactions, pairwise comparisons based on the difference of the estimated means were performed using the lsmeans package (Lenth, 2016); in order to account for planned multiple comparison, a Benjamini–Hochberg procedure was used.

Results

Because of the inherent room acoustic characteristics of the laboratory classrooms, different listening conditions were realized for the 5Y on the one hand and the 6Y and 7Y children on the other hand. For this reason, the two groups could not be directly compared, and their results were analyzed separately.

Effects of Room Acoustics on Speech Reception: 5-Year-Old Pupils

In the setup of the statistical models, listening condition, test repetition, and the interaction of these two variables were considered as fixed factors, and the child was considered as a random factor. Thus, the response variable data were not averaged across participants prior to analysis, but the individual responses to the fixed factors of the experiment were estimated for each participant by the GLMM model. Only one class of the 5Y children performed the experiment, and the listening conditions were presented in the order A (quiet–working–quiet–working classroom).

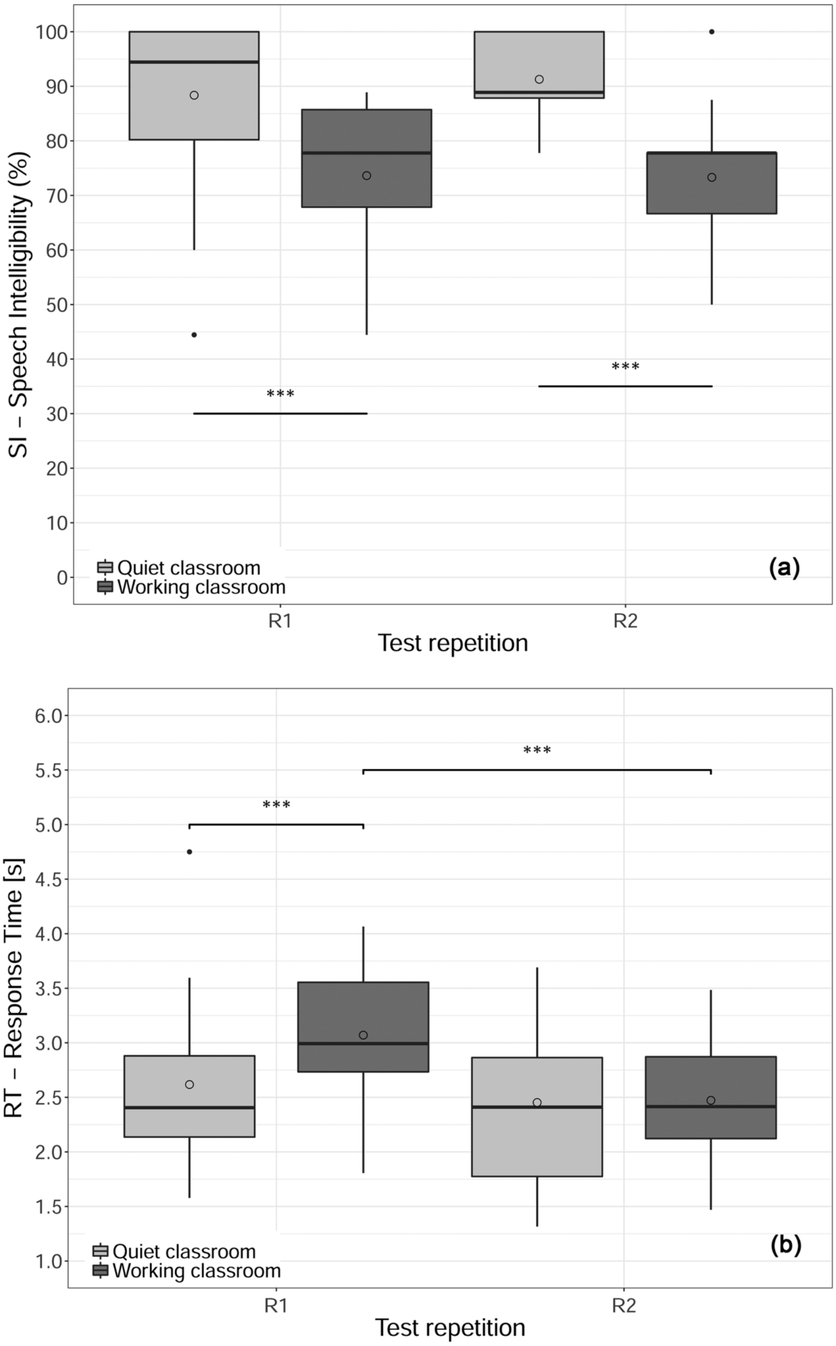
The SI results for the listening conditions and the test repetitions are shown in Figure 3a. The statistical model **F3** revealed that the effect of listening condition was significant, $\chi^2(1) = 29.8$, $p < .001$. The test repetition had no significant influence on the SI results ($p = .45$), and neither had the interaction of the two variables ($p = .57$). Pairwise comparisons of the SI results averaged across the repetitions showed that the participants scored significantly better in the "quiet classroom" as compared to the "working classroom" condition, with a difference between the SI mean values of 16.5% (quiet classroom: 90.7%; working classroom: 74.2%).

Figure 3b details the results of the 5Y children with regard to the RT data. For the analysis, all responses (correct and incorrect) were used; considering the latencies of correct responses alone did not alter the statistical outcomes. The statistical analysis showed a significant effect of the listening condition, $\chi^2(1) = 8.64$, $p = .003$; the test repetition, $\chi^2(1) = 13.56$, $p < .001$; and their interaction, $\chi^2(1) = 3.97$, $p = .046$. The presence of an interaction between the two factors indicated that the RT results varied differently over time, depending on the listening condition. In particular, when analyzing the pairwise comparisons across the test repetitions, no difference was found between the RTs of R1 and R2 ($p = .22$) for the "quiet classroom" condition. Differently, when the background noise was played back, the RTs changed significantly over the repetitions (R1: 3,044 ms, R2: 2,455 ms; $z = 4.00$, $p < .001$), with an estimated mean decrease of 589 ms. Focusing on the comparison between the listening conditions, the post hoc tests indicated thepresence of a significant increase in RT of 522 ms in the "working classroom" condition versus the "quiet classroom" condition in the first repetition alone (quiet classroom: 2522 ms; working classroom: 3044 ms; $z = -3.48$, $p < .001$). In R2, the RTs did not significantly differ between the listening conditions ($p = .48$).

Effects of Room Acoustics on Speech Reception: Primary School Pupils (6–7 Years Old)

The results of primary school children were analyzed using statistical models with listening condition, test repetition, age (6Y vs. 7Y), and condition order as fixed factors and the child as a random factor. The two-way interactions between the factors were also considered in the models.

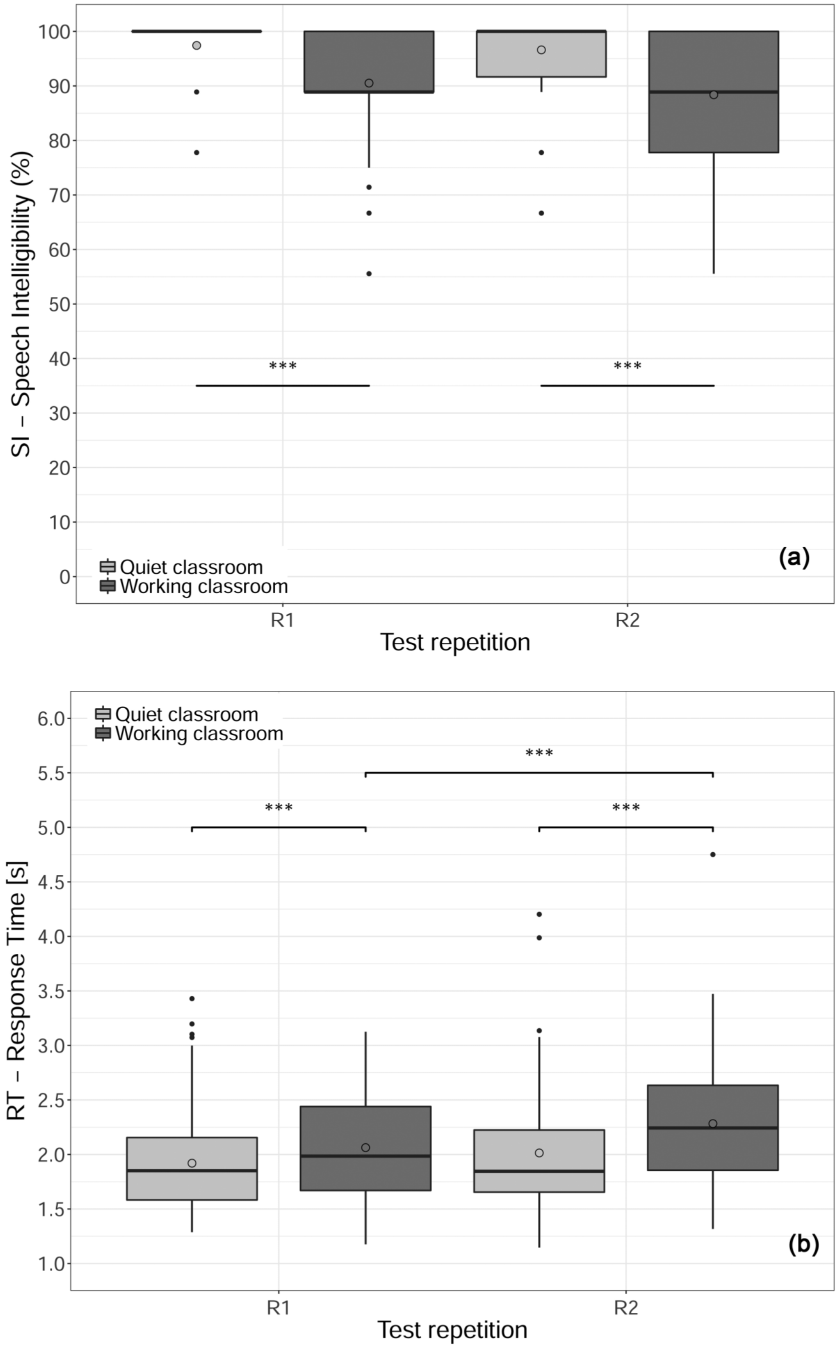
Figure 3. Box plots of the (a) speech intelligibility and (b) response time results for kindergarten (5Y) children. The results are organized according to the listening condition (quiet vs. working classroom) and the test repetition (R1 vs. R2). The branches represent pairwise significant differences, as found by the post hoc analysis (*** corresponds to $p < .001$). The bottom and the top of the boxes are the first and the third quartiles of the AQ6 data distributions, respectively; the central bold lines are the median values, and the circles represent the mean values; 99% of the data lay within the whiskers. The outliers are shown as points outside the whiskers.



The SI results, averaged over age and condition order, F4 are presented in Figure 4a. The analysis indicated a significant main effect of listening condition alone, $\chi^2(1) = 59.45$, $p < .001$. The effects of age ($p = .46$), test repetition ($p = .10$), condition order ($p = .12$), and their interactions were not statistically significant. Post hoc pairwise comparisons between the two listening conditions revealed that the average accuracy in the “quiet classroom” condition (SI = 97.3%)

AQ7

Figure 4. Box plots of (a) speech intelligibility and (b) response time results for primary school children (6Y and 7Y). The results are organized according to the listening condition (quiet vs. working classroom) and the test repetition (R1 vs. R2). The branches represent pairwise significant differences, as found by the post hoc analysis (*** corresponds to $p < .001$). The bottom and the top of the boxes are the first and third quartiles of the data distributions, respectively; the central bold lines are the median values, and the circles represent the mean values; 99% of the data lay within the whiskers. The outliers are shown as points outside the whiskers.



was significantly higher than the accuracy obtained in the “working classroom” condition (SI = 90.2%), with a mean difference of 7.1%.

With regard to the RT results, the descriptive statistics are depicted in Figure 4b. The GLMM model showed that the age of the participants was not significant ($p = .76$) nor was the condition order ($p = .44$). However, listening condition, $\chi^2(1) = 48.70$, $p < .001$; test repetition, $\chi^2(1) = 21.86$, $p < .001$; and their interaction, $\chi^2(1) = 4.10$, $p = .043$, did have a significant effect. None of the other interactions reached the level of significance. When examining the pairwise comparisons between listening conditions for each test repetition, it was found that they were always significant (R1—quiet classroom: 1,885 ms, working classroom: 2,026 ms, $z = -3.56$, $p < .001$; R2—quiet classroom: 1,959 ms, working classroom: 2,231 ms, $z = -4.74$, $p < .001$), indicating that longer RTs were measured in the presence of added background noise. It is worth noting that, as a result of the significant interaction between test repetition and condition, the estimated increase in RT due to the presence of noise almost doubles in the second test repetition (R1: ΔRT = 141 ms; R2: ΔRT = 272 ms). Then, when considering the pairwise RT comparison over the repetitions, a significant increase was found in the "working classroom" condition alone. The result indicates that when the background noise was played back, the children were significantly slower in the second test repetition (R1: 2,026 ms, R2: 2,231 ms; ΔRT = 205 ms, $z = -4.74$, $p < .001$). No significant difference was found in the "quiet classroom" condition ($p = .058$).

Discussion

Markers of Listening Effort: 5-Year-Old Pupils

The first aim of this work was to assess the effect of acoustic conditions and test repetition on both accuracy and RT. The measures were compared from the "quiet" to the "working classroom" condition and from the first to the second repetition, and a slowing down in the RT results was taken as an indicator of a heavier processing load.

With regard to the group of the 5Y pupils, the statistical analysis returned a significant main effect of listening condition for SI, indicating, as expected, a large decrease in task accuracy when noise was played back. The SI results were maintained over the two test repetitions, indicating that the children continued to carry out the task throughout the experiment, without quitting the activity, and that no learning effects were observed in the accuracy of the results. Differently, a significant interaction between the listening condition and test repetition was obtained for RT, pointing out that the expected increase in cognitive load in the presence of added background noise was only realized in the R1 repetition (ΔRT = 522 ms). In the second repetition (R2), the RT values did not differ in the two conditions. Therefore, although the children suffered, in both test repetitions, from a large decrease in intelligibility scores due to the presence of noise, in R2 they responded faster than in R1 when noise was played back and as quickly as in the "quiet classroom" condition.

It can be hypothesized that the RT decrease in noisy R2 was driven by a learning effect, developing over the course of the experiment for the more challenging condition and yielding an improvement in the RT results with continued practice. The learning effect was not apparent in the "quiet classroom" condition and did not imply an increase in SI results in either case. However, in turn, this may be due to the children having reached, at least in the "quiet classroom," their optimal performance for both RT and SI (albeit not 100% SI) and hence there being no room for further improvement.

More research is needed to understand the interaction between noise and repetition for 5Y children, aiming to assess the feasibility of the single-task paradigm for the youngest pupils. Indeed, if the hypothesis of a learning effect was accurate, additional practice should be included in the experimental paradigm, while still complying with the constraint of a short test duration.

Markers of Listening Effort: 6- and 7-Year-Old Pupils

With regard to the 6Y and 7Y, no age effect was outlined in the statistical model, neither for SI nor RT. This result differs from previous findings. In fact, Prodi et al. (2013) tested a panel of 55 6Y and 156 7Y in classrooms using the same equipment, procedure, and speech material as in the present case, though the noise employed in that study resembled traffic, activity, and tapping. Especially in acoustical conditions comparable to the present case (i.e., similar STI values), the 7Y displayed higher SI results and faster RTs than the 6Y, and the gap was significant in the quiet condition as well. It is known from studies on the impact of noise of serial recall that noise characteristics elicit peculiar effects on children. In particular, noise with a changing state has a direct access to short-term memory, as occurs for adults (Hughes & Jones, 2001), and this mechanism is not much affected by developmental factors. On the contrary, noise attention capture potentials are more effective for younger children due to their lesser ability to resist distraction (Klatte et al., 2013; Klatte, Lachmann, Schlittmeier, & Hellbrück, 2010). The noises used by Prodi et al. (2013) had both a changing state and a degree of salience that helped in stretching the performance gap due to age. On the other hand, the present LTSS noise had at least such characteristics, and for this reason, the absence of age effects between the 6Y and 7Y was not unexpected. Furthermore, Prodi et al. (2013) did not include a baseline condition, and the comparison of the three noise types yielded a partial ranking for the 6Y and 7Y, with traffic noise producing longer latencies in some cases. The present results show that, disregarding the lesser disrupting potential of LTSS, the presence of noise always causes lower SI and longer RT with respect to a baseline in both repetitions. This occurrence is consistent with the results of Leibold et al. (2016) and witnesses an increased involvement of cognitive resources even with noise having basic spectrotemporal features (broadband, continuous, and steady state). The observed slowing down of the RTs in more adverse listening conditions can thus be interpreted as a slowing down of the speech processing rate (McGarrigle et al., 2014) and support the feasibility of using a single-task paradigm to measure the listening effort for 6- and 7-year-old pupils. The present finding adds to the results of Lewis et al. (2016), who found that a worsening in SNR was paired with both a significant decrease in SI results and a significant increase in onset RTs for 5- to 12-year-old pupils. Furthermore, it expands the outcomes of a study of Prodi, Visentin, and Farnetani (2010) where the link between SI, RT, and acoustical conditions (room treatment and reverberation) was outlined for pupils aged 8–10 years; the study results indicated a significant decrease in SI and a significant slowing down of RT as the classroom acoustics worsened.

Then, considering the 6Y and 7Y together from now on, the comparison between the SI and RT data across the repetitions is the most relevant in the results. First, it can be seen that, for each listening condition, the intelligibility scores in R1 and R2 were not significantly different, even though in R2 more variability in the SI results was observed; children were thus able to mobilize capacity in order to keep the same accuracy along the test duration. On the contrary, the trend of RT data shows that, while in the "quiet classroom" condition, similar RTs were measured in R1 and R2; when background noise was added, a slowing down of the RTs occurred in R2. This finding is consistent with an increase in the pupils' listening effort during the monitored interval, and the effect is thought to be driven primarily by the presence of the background noise. Indeed, the result could be potentially explained by additional causes to noise alone, for instance, boredom or distraction elicited by the duration of the task. Nevertheless, the statistical analysis of the RT data returned no significant effect or interactions for the presentation order of the listening condition, thus indicating that effects related to the task duration could be ruled out and the additional slowing down of the RTs in R2 could be explained by the effect of the background noise. A similar trend in the RT results over the course of a 45-min period was outlined by Prodi and Visentin (2015), with reference to older children (8–10 years of age) and favorable listening conditions (STI corresponding to "fair" or "good" intelligibility).

It is interesting to notice the difference between the 5Y children and the 6Y and 7Y children concerning the effect of test repetition on RT in the "working classroom" condition. Indeed, the youngest children showed faster RTs in R2, whereas the older children showed slower RTs. Several hypotheses might be formulated, related either to the difference in the acoustical conditions (i.e., the 5Y performed the experiment in more challenging conditions; see Table 2) or the presence of learning effects for the youngest pupils or coping mechanisms developing in the examined age range. Dedicated experiments are needed to establish why this difference exists.

Role of Room Acoustics in Effortful Speech Reception

Finally, with regard to the second research question of the study, it was possible to better outline the role that the room acoustics has in the insurgence of listening effort. The SNR and STI values of the "quiet classroom" condition that kept the RT results unaffected during repetitions were congruent with the earlier recommendations of SNR $\geq +15$ dB (Sato & Bradley, 2008) and with the normative Italian recommendations of STI ≥ 0.6 (Ente Nazionale Italiano di Unificazione, 2010). Unfortunately, such values are not easily ensured during regular educational activities; the acoustic conditions often depart from the optimal conditions with an increase in the background noise level up to 20 dB(A), depending on the pupils' activity during the school day (Shield & Dockrell, 2004). If in a given room the SNR departs from the optimal value, the present findings indicate that, together with a decrease in the accuracy of the speech reception performance, an increase in the RTs also occurs, indicating that more effort is expended by the older children. In these conditions, in order to maintain the same accuracy over the lesson, the 6Y and 7Y children deploy progressively greater cognitive resources. Supposedly, an increase in RTs already in speech reception will also impair the speech communication, affecting meaning extraction and recall. However, how these effects reflect on more specific academic tasks and on the children's learning and academic achievements should be investigated in specific experiments.

When relying on the present data, it is not possible to set a precise value in terms of objective indicators that warrant the absence of RT changes during lessons. Nevertheless, it cannot be excluded that the present normative limits are unsatisfactory to avoid the drift of RT and hence listening effort during a noisy lesson.

Limitations

The experimental paradigm relied on a community presentation of a speech reception task in the familiar context of the classroom, with the children surrounded by their classmates. This is a natural and unconstrained setting for young children, which ensures an ecologically valid environmental situation and should preserve the children's spatial awareness. However, some aspects of the present approach may limit its ecological validity and need to be addressed in future works. First, the experiment was based on a speech reception task, which is not directly representative of the cognitive processes relied upon during typical classroom activities (e.g., sentence recall, speech comprehension, and extraction of message). However, the correct (and easy) identification of what the teacher said is a necessary condition subtending all the subsequent higher order processes (Hygge, 2014). Then, even though the task choice yielded moving away from a fully ecological paradigm, it nevertheless warranted generality to the approach. In the future, experiments with more specific and demanding tasks, better representing the academic activities of young pupils, could be employed to outline the role of acoustic conditions on memory and learning performance. Presenting a task in which the pupils can identify or consider as meaningful or familiar may also stimulate the children to commit their mental resources to the task and then to potentially obtain a better match between motivation during the experiment and everyday activities. Second, the speech signal was played back through a loudspeaker in order to ensure a highly controlled and reproducible target signal. Nevertheless, the audio-only presentation was not able to provide the children with all the nonaural cues (i.e., visual and gestural), which may assist or distract the performance in a real-life listening situation. Then, aiming at a better ecological validity of the experimental paradigm, the effect of multisensory integration should also be accounted for.

Conclusions

In this study, the effects of classroom acoustics on the speech reception of 5- to 7-year-old pupils were investigated using an ecological experimental approach. With reference to the research questions motivating the experiment, the following results were found:

1. The single-task RT paradigm was effective in tracing changes in the processing effort of the older children. In fact, the RTs significantly slowed down when a stationary noise was added, witnessing an increased involvement of processing resources compared to a baseline condition. The finding is relevant as it demonstrates that the metric can be successfully employed for 6Y and 7Y children to track changes in the acoustic conditions.
2. The monitoring of the RTs during a time interval for 5Y kindergarten pupils highlighted a peculiar trend. In the presence of stationary background noise, their RT results decreased upon test repetition. More research is needed to establish whether the result is yielded by the need of additional practice in the task.
3. The trends of the RT results of the 6Y and 7Y pupils over the test repetition indicated that, in order to maintain the same performance accuracy in background noise, the children needed to deploy progressively more cognitive resources. The result points toward a change in the listening effort during a lesson period, appearing very plausible in most of their daily working conditions.
4. Only when the normative classroom requirements were met (i.e., SNR > +15 dB, corresponding to the children being involved in isolated and quiet activities), the RTs were found to be constant over the repetitions. Moving away from this favorable condition, due to children engaged in noisier activities, yielded a negative effect on both performance and listening effort, the latter changing over the repetitions.

Overall, the results of this study point toward the effectiveness of the experimental paradigm in tracing the 6Y and 7Y children's listening effort. The finding is valuable and has implications in terms of an enhanced acoustic design of the classrooms, best tailored to the children's needs. On this issue, Visentin, Prodi, Cappelletti, Torresin, and Gasparella (2018) showed that, with regard to young adults with normal hearing, realistic modifications to the acoustic conditions of the room yielding similar accuracy in a word recognition task were found to change the amount of processing resources involved, the latter monitored by using the single-task RT. The experimental paradigm presented in this study, assessing children's accuracy and listening effort simultaneously, shows then potential in addressing the impact of an ecological classroom acoustic environment (and its modifications) on children with normal hearing. Further work is needed to assess its effectiveness for children with hearing impairment or who wear hearing aids.

Acknowledgments

AQ8

Aspects of this work were presented at the Euronoise 2018 Conference. The authors wish to thank the teachers and the pupils involved in the experiment for their participation.

References

Anderson, K. (2004). The problem of classroom acoustics: The typical classroom soundscape is a barrier to learning. *Seminars in Hearing, 25*(2), 117–129.

Arslan, E., Genovese, E., Orzan, E., & Turrini, M. (1997). Valutazione della percezione verbale nel bambino ipoacusico: Test di Identificazione di Parole Infantili 1–2 (Evaluation of verbal perception in children with hearing loss: Infant Words Identification Test 1–2). Bari, Italy: Ecumenica.

Astolfi, A., Bottalico, P., & Barbato, G. (2012). Subjective and objective speech intelligibility investigations in primary school classrooms. *The Journal of the Acoustical Society of America, 131*(1), 247–257.

Baayen, H. R., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research, 3*(2), 12–28.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed effects models using lme4. *Journal of Statistical Software, 67,* 1–48.

Bess, F. H., & Hornsby, B. W. (2014). Commentary: Listening can be exhausting—Fatigue in children and adults with hearing loss. *Ear and Hearing, 35*(6), 592–599.

Bottalico, P., & Astolfi, A. (2012). Investigations into vocal doses and parameters pertaining to primary school teachers in classrooms. *The Journal of the Acoustical Society of America, 131*(4), 2817–2827.

Choi, S., Lotto, A., Lewis, D., Hoover, B., & Stelmachowicz, P. (2008). Attentional modulation of word recognition by children in a dual-task paradigm. *Journal of Speech, Language, and Hearing Research, 51*(4), 1042–1054.

Cowan, N., Towse, J. N., Hamilton, Z., Saults, J. S., Elliott, E. M., Lacey, J. F., . . . Hitch, G. J. (2003). Children's working-memory processes: A response-timing analysis. *Journal of Experimental Psychology: General, 132*(1), 113–132.

Crandell, C. C., & Smaldino, J. J. (2000). Classroom acoustics for children with normal hearing and with hearing impairment. *Language, Speech, and Hearing Services in Schools, 31*(4), 362–370.

Ente Nazionale Italiano di Unificazione. (2010). Acustica in edilizia—Classificazione acustica delle unità immobiliari, UNI 11367:2010–Annesso A *[Building acoustics—Acoustical classification of real estate, UNI 11367:2010–Annex A)*. Milan, Italy. AQ9

Everitt, B. S., & Hothorn, T. (2010). *A handbook of statistical analysis using R* (2nd ed.). New York, NY: Chapman and Hall/CRC.

Gagné, J. P., Besser, J., & Lemke, U. (2017). Behavioral assessment of listening effort using a dual-task paradigm: A review. *Trends in Hearing, 21,* 1–25.

Gustafson, S., McCreery, R., Hoover, B., Kopun, J. G., & Stelmachowicz, P. (2014). Listening effort and perceived clarity for normal hearing children with the use of digital noise reduction. *Ear and Hearing, 35*(2), 183–194.

Hicks, C. B., & Tharpe, A. M. (2002). Listening effort and fatigue in school-age children with and without hearing loss. *Journal of Speech, Language, and Hearing Research, 45*(3), 573–584.

Houben, R., van Doorn-Bierman, M., & Dreschler, W. A. (2013). Using response time to speech as a measure for listening effort. *International Journal of Audiology, 52*(11), 753–761.

Howard, C. S., Munro, K. J., & Plack, C. J. (2010). Listening effort at signal-to-noise ratios that are typical of the school classroom. *International Journal of Audiology, 49*(12), 928–932.

Hughes, R., & Jones, D. M. (2001). The intrusiveness of sound: Laboratory findings and their implications for noise abatement. *Noise Health, 4*(13), 51–70.

Hygge, S. (2014). Classroom noise and its effect on learning. *Proceedings of the 11th International Congress on Noise as a Public Health Problem (ICBEN), Nara, Japan* (Vol. 4). Retrieved from http://hig.diva-portal.org/smash/record.jsf?pid=diva2%3A731143&dswid=3671

International Electrotechnical Commission. (2011). *Sound system equipment. Part 16: Objective rating of speech intelligibility by speech transmission index: IEC 60286-16*. Geneva, Switzerland: Author.

International Organization of Standardization. (2003). *Ergonomics—Assessment of speech communication ISO9921:2003*. Geneva, Switzerland: Author.

International Organization of Standardization. (2008). *Acoustics—Measurement of room acoustic parameters. Part 2: Reverberation time in ordinary rooms ISO3382-2:2008*. Geneva, Switzerland: Author.

International Organization of Standardization. (2009). Acoustics—Measurement of room acoustic parameters. In *Part 1: Performance spaces ISO3382-1:2009*. Geneva, Switzerland: Author.

Kahnemann, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.

Klatte, M., Bergström, K., & Lachmann, T. (2013). Does noise affect learning? A short review on noise effects on cognitive performance in children. *Frontiers in Psychology, 4,* 578.

Klatte, M., Lachmann, T., Schlittmeier, S., & Hellbrück, J. (2010). The irrelevant sound effect in short-term memory: Is there developmental change? *European Journal of Cognitive Psychology, 22*(8), 1168–1191.

Klink, K. B., Schulte, M., & Meis, M. (2012a). Measuring listening effort in the field of audiology—A literature review of methods (Part 1). *Zeitschrift für Audiologie, 51*(2), 60–67.

Klink, K. B., Schulte, M., & Meis, M. (2012b). Measuring Listening effort in the field of audiology—A literature review of methods (Part 2). *Zeitschrift für Audiologie, 51*(3), 96–105.

Leibold, L., Bonino, A. Y., & Buss, E. (2016). Masked speech perception in infants children and adults. *Ear and Hearing, 37*(3), 345–353.

Lenth, R. V. (2016). Least-squares means: The R package. *Journal of Statistical Software, 69*(1), 1–33.

Lewis, D., Schmid, K., O'Leary, S., Spalding, J., Heinrichs-Graham, E., & High, R. (2016). Effects of noise on speech recognition and listening effort in children with normal hearing and children with mild bilateral or unilateral hearing loss. *Journal of Speech, Language, and Hearing Research, 59*(5), 1218–1232.

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyze reaction time data. *Frontiers in Psychology, 6,* 1171.

McCreery, R. W., & Stelmachowicz, P. G. (2013). The effects of limited bandwidth and noise on verbal processing time and word recall in normal-hearing children. *Ear and Hearing, 34*(5), 585–591.

McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., & Amitay, S. (2014). Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper.' *International Journal of Audiology, 53*(7), 433–440.

Mealings, K. T., Buchholz, J. M., Demuth, K., & Dillon, H. (2015). Investigating the acoustics of a sample of open plan and enclosed Kindergarten classrooms in Australia. *Applied Acoustics, 100,* 95–105.

Nelson, P. B., & Soli, S. (2000). Acoustical barriers to learning: Children at risk in every classroom. *Language, Speech, and Hearing Services in Schools, 31*(4), 356–361.

Pals, C., Sarampalis, A., van Rijn, H., & Başkent, D. (2015). Validation of a simple response-time measure of listening effort. *The Journal of the Acoustical Society of America, 138*(3), EL187–EL192.

Picard, M., & Bradley, J. S. (2001). Revisiting speech interference in classrooms. *Audiology, 40*(5), 221–244.

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., . . . Wingfield, A. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear and Hearing, 37* (Suppl. 1), 5S–27S.

Prodi, N., & Visentin, C. (2015). Listening efficiency during lessons under various types of noise. *The Journal of the Acoustical Society of America, 138*(4), 2438–2448.

Prodi, N., Visentin, C., & Bellettini, C. (2012, March). Listening efficiency testing. *Paper presented at the 45th International Conference: Applications of Time–Frequency Processing in Audio, Ferrara, Italy*.

Prodi, N., Visentin, C., & Farnetani, A. (2010). Intelligibility, listening difficulty and listening efficiency in auralized classrooms. *The Journal of the Acoustical Society of America, 128*(1), 172–181.

Prodi, N., Visentin, C., & Feletti, A. (2013). On the perception of speech in primary school classrooms: Ranking of noise interference and of age influence. *The Journal of the Acoustical Society of America, 133*(1), 255–268.

Reis, H. T. (2012). Why researchers should think "real-world": A conceptual rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life*. New York, NY: Guilford.

Sato, H., & Bradley, J. S. (2008). Evaluation of acoustical conditions for speech communication in working elementary school classrooms. *The Journal of the Acoustical Society of America, 123*(4), 2064–2077.

Shield, B. M., & Dockrell, J. E. (2003). The effects of noise on children at school: A review. *Building Acoustics, 10*(2), 97–116.

Shield, B. M., & Dockrell, J. E. (2004). External and internal noise surveys of London primary schools. *The Journal of the Acoustical Society of America, 115*(2), 730–738.

Shield, B. M., & Dockrell, J. E. (2008). The effects of environmental and classroom noise on the academic attainments of primary school children. *The Journal of the Acoustical Society of America, 123*(1), 133–144.

Valente, D. L., Plevinsky, H. M., Franco, J. M., Heinrichs-Graham, E. C., & Lewis, D. E. (2012). Experimental investigation of the effects of the acoustical conditions in a simulated classroom on speech recognition and learning in children. *The Journal of the Acoustical Society of America, 131*(1), 232–246.

van den Tillan-Haverkate, M., de Ronde-Brons, I., Dreschler, W. A., & Houben, R. (2017). The influence of noise reduction on speech intelligibility, response times to speech, and perceived listening effort in normal-hearing listeners. *Trends in Hearing, 21,* 1–13.

Vickers, D. A., Backus, B. C., Macdonald, N. K., Rostamzadeh, N. K., Mason, N. K., Pandya, R., Mahon, M. H. (2013). Using personal response systems to assess speech perception within the classroom: An approach to determine the efficacy of sound field amplification in primary school classrooms. *Ear and Hearing, 34*(4), 491–502.

Visentin, C., Prodi, N., Cappelletti, F., Torresin, S., & Gasparella, A. (2018). Using listening effort assessment in the acoustical design of rooms for speech. *Building and Environment, 136,* 38–53.

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record, 58*(3), 475–482.

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

AQ1: This article has been edited for grammar, APA style, and usage. Please use annotations to make corrections on this PDF. Please limit your corrections to substantive changes that affect meaning. If no change is required in response to a question, please respond "OK as set."

AQ2: Please confirm whether the corresponding author should be Dr. Prodi or Dr. Visentin.

AQ3: Choi, Lotto, Lewis, Hoover and Stelmachowicz (2016) is missing from the reference list. Please provide reference details or remove the reference from the text.

AQ4: Please spell out TIPI here.

AQ5: R Core Team, 2017, is missing from the reference list. Please provide reference details or remove the reference from the text.

AQ6: Please check if "respectively" is appropriate here.

AQ7: Please check if "respectively" is appropriate here.

AQ8: Authors typically use an Acknowledgments section to specify grant or funding information. If applicable, please add this information here.

AQ9: Kindly provide publisher name.

END OF ALL QUERIES